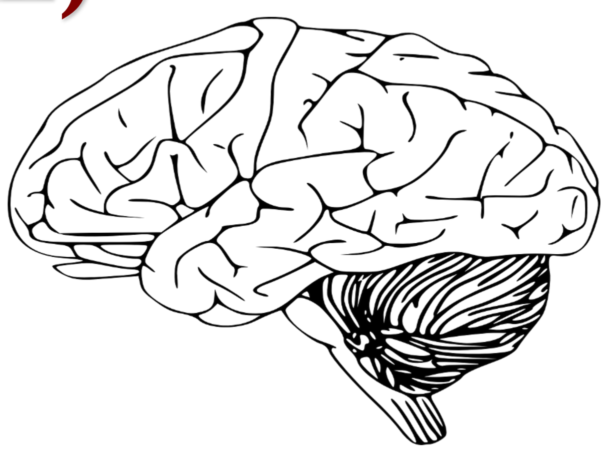
BRanching Artificial Neural Ensemble (BRANE) Algorithm for Supervised Learning

Muhammad Wasimuddin, Viyaleta Peterson, Karan Bhosale
Department of Computer Science
University of Bridgeport, Bridgeport, CT



Abstract

Various models exist to predict a numerical value in supervised learning problems. One of the challenges in predicting an outcome with high degree of precision involves dealing with numerical data points which can be represented using differently. To solve for such challenge and in order to predict the *logerror* value in Zillow's competition on Kaggle, we have developed a new model, **BR**anching **A**rtificial **N**eural **E**nsemble (BRANE). This ensemble network uses a number of multilayer perceptrons (MLP) to predict the outcome and combines the results using an additional MLP. This approach not only allowed us to use different datatypes as inputs, but also predicted better and converged faster than traditional MLP models.

Problem Definition and Dataset

Problem Definition: the error in house transaction needs to be predicted in logarithmic form based on the provided log error for previous house transactions.

logerror = log(*Zestimate*) – log(*SalesPrice*)

Dataset: the dataset provided in the competition contains 2.9M houses with 58 attributes and 90,275 sale transactions of houses in 2016.

Preprocessing: the transaction dataset consists of *parcelid* and *logerror*. The houses dataset consists of the house attributes. The two datasets undergo reconciliation before preprocessing so that the training dataset consist of attribute values and log error.

- Drop all the attributes that have more than 50 percent of their data missing (Figure 1).
- Generate a correlation matrix. Find the highly co related attributes and trop them (Figure 2).
- With eleven attributes left create a numerical dataset by filling the missing values with mean of each attribute
- Create a second, binary dataset, by flagging the missing values with "1" and "0" for non missing values

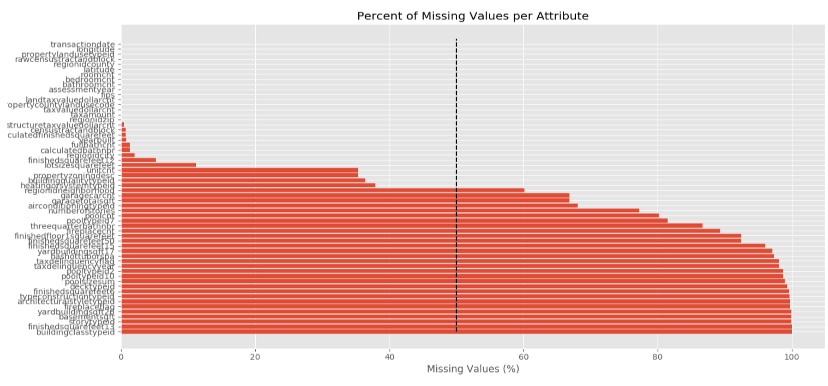


Figure 1

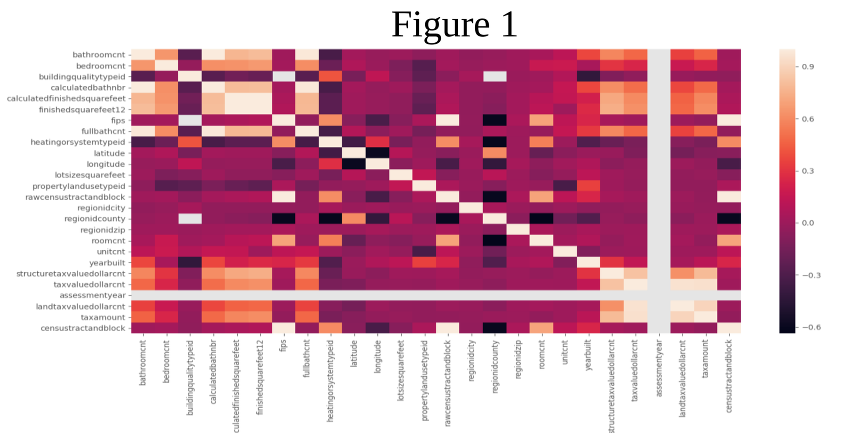


Figure 2

Methodology

BRanching **A**rtificial **N**eural **E**nsemble (Figure 3) is comprised of three MLPs. The first sensor MLP predicts the *logerror* using the numerical values of the attributes. The second sensor MLP predicts *logerror* using binary values of the attributes. The decision MLP uses the outputs of the two sensor MLPs as inputs. The final prediction is provided by the output of decision MLP. The MLPs have their own back propagation mechanism and the error from the decision MLP is not backpropagated to the sensor MLPs.

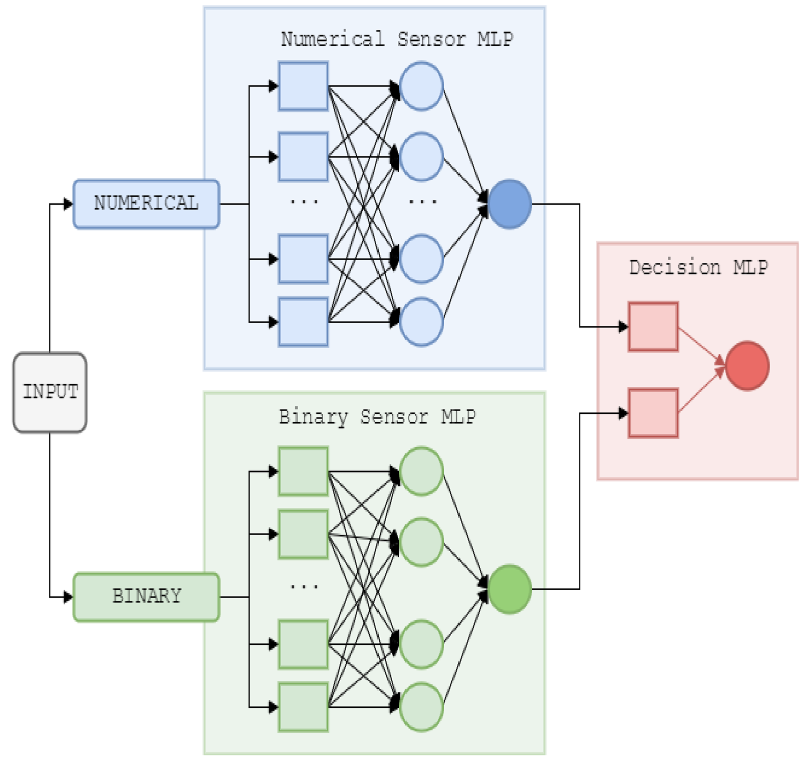


Figure 3

Forward and Back Propagation.
In feed forward propagation, the weights are randomly assigned from uniform range [0, 1]. In hidden layers, sigmoid activation function is applied in the sensor MLPs and linear activation function is applied in decision MLP.

Forward propagation:
Input values of each MLP are fed forward via the following equation, where x_j is neuron at current layer, x_i is neuron value at previous layer, w_{ij} is the weight of the edge between the two neurons, b_i is bias from previous layer.

$$x_j = f\ (\Sigma\, w_{ij} \cdot x_i\ +\ b_i)$$

Backward propagation:
Mean Square Error (MSE) loss function is used in gradient descent optimization, where $\hat{y}_i$ is the predicted value and y_i is the ground truth:

$$MSE = (\hat{y}_i - y_i)^2/2$$

The weights are optimized through back propagation using δ - partial derivatives form output layers, η – the learning rate, w_{ij} - weight before backpropagation, and x_j - neuron value.

$$w_{ij} = w_{ij} + \eta \cdot \delta \cdot x_j$$

Results and Evaluation

We performed experiments on small dataset that was to test the effectiveness of the dataset. First two experiments use a small batch of 600 and 300 records for training and testing respectively. The third experiment was carried out on the entire dataset.

Experiment 1
In the first experiment we compared the performance between the BRANE Algorithm versus binary dataset using one standard MLP with one hidden layer versus numerical dataset using one standard MLP and one hidden layer. (Figure 4)

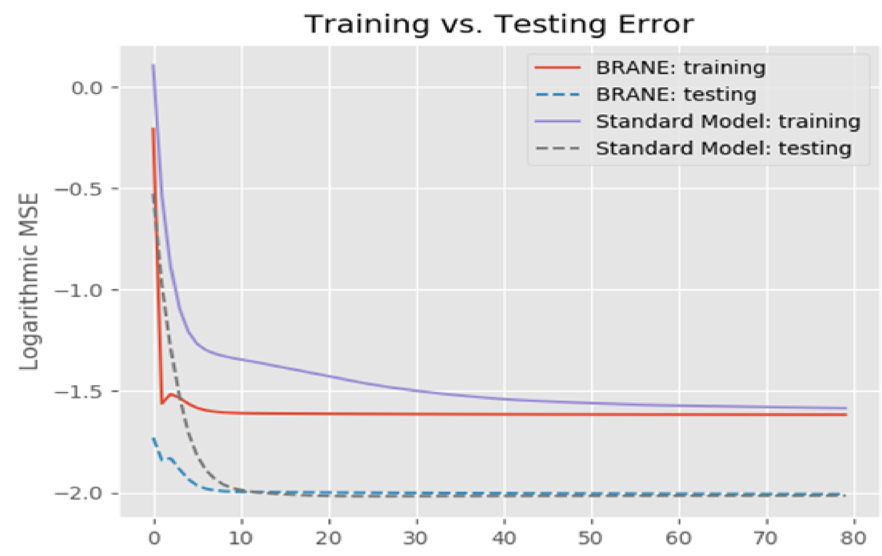


Figure 4

Experiment 2.
The second experiment was on a small batch to compare BRANE architecture errors versus the errors obtained from numerical sensor and binary sensor individually. (Figure 5)

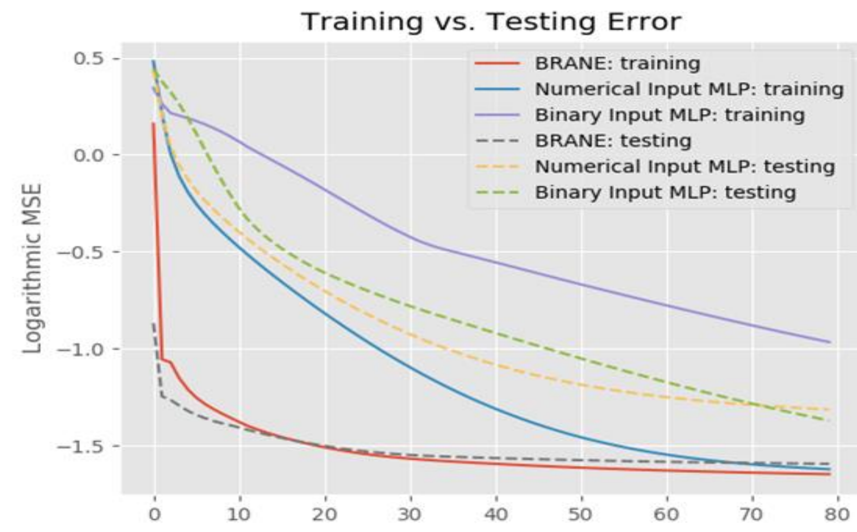


Figure 5

Experiment 3.
Finally we carried out the experiment on the entire dataset instead of small batch. A learning rate of 0.001 and 50 epoch were used to train algorithm. The figure 6. shows that the BRANE algorithm converges faster and has a low MSE error. Table 1. shows the MSE comparison for training and testing dataset, clearly showing low MSE for BRANE

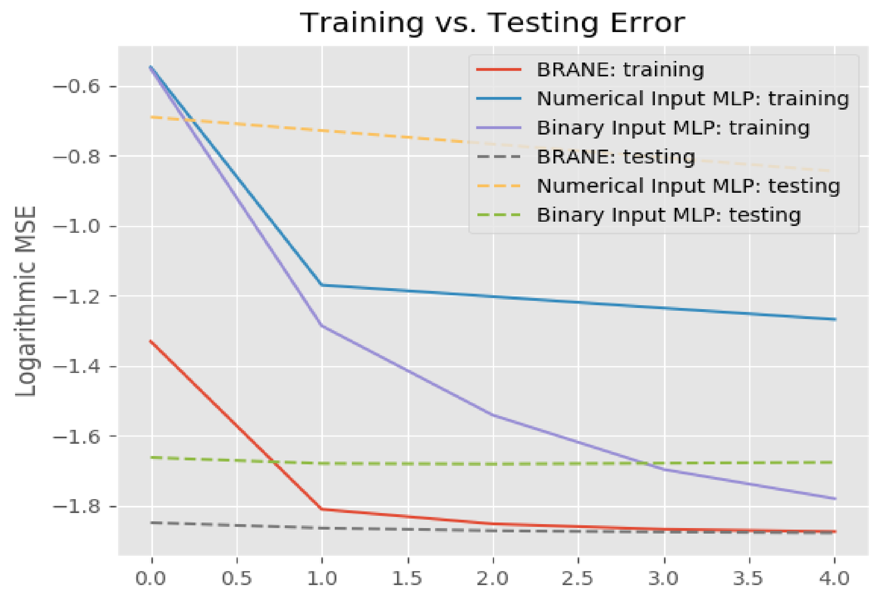


Figure 6

Phase	Algorithm	MSE
Training	Numerical MLP	0.014172
	Binary MLP	0.013695
	BRANE	0.013631
Testing	Numerical MLP	0.012948
	Binary MLP	0.012096
	BRANE	0.011748

Table 1.

Accuracy - The training MSE of 1.36% and testing MSE of 1.18%. Since the error is so low, we consider that our model provided accurate results. Furthermore, the BRANE algorithm scored better results than the outcomes of each sensor MLP, when considered individually.
Speed - In our analysis, we found that BRANE reduces computation complexity of feedforward alone by about 30%.
Robustness - Robustness is the measure of how fast the network converges. It is evident from our experiments that BRANE converges faster than other MLPs at a lower learning rate.
Scalability - The BRANE algorithm is scalable in terms of number of sensors, number of features and number of hidden layers.
Interpretability - The output value predicts what the problem asked for: the *logerror*, indicating how well Zestimate is able to predict home values.

Conclusion

Human Brain uses multiple senses to properly identify objects. Based on this logic we have built a neural network BRANE by Ensembling multiple sensor MLP network and feeding the output to decision MLP to predict the accuracy of Zillow's Zestimate Algorithm. With our approach we have obtained error of 1.18% and 9.27% error reduction when compared to numerical sensor MLP, and 2.88 reduction when compared to binary MLP. BRANE showed a significant time complexity improvement from the standard MLP with twice as many inputs. Finally, BRANE converges 135% faster than the stand-alone Numerical sensor MLP and 25.9% faster than the stand-alone Binary sensor MLP, making it more efficient in predicting the accuracy.

Dr. Jeongkyu Lee
CPSC-552-6T1-Data Mining-2017FA